

Principal Components of a Multispectral Image: Application to a Geological Problem

Abstract: The principal components of a multispectral image are studied with special emphasis placed on photointerpretation applications. The method used to obtain these components involves a linear radiometric enhancement which avoids artifact creation and does not require extra computational work. The results have been particularized to Landsat multispectral scanner images and applied to the geological study of a sedimentary basin in central Spain.

Introduction

Multispectral remote sensors provide data in the form of several spectral images of the observed object. Each image represents the spatial distribution of reflected light of frequencies lying in the corresponding spectral window. Therefore, rather high correlation is expected between these different spectral images of the same object. The actual degree of correlation depends on the physical nature of the object, the widths of the spectral windows, and the amount and nature of the noise present in the data. Because of these factors, multispectral digital images contain a certain amount of data redundancy; however, this can be reduced by choosing the less correlated bands. For instance, to make color prints (where at most three bands can be assigned to the fundamental colors) the three less correlated bands can be used. In this manner some redundant data are retained while a certain amount of information present only in the omitted bands is lost. A similar situation occurs when one supplies less correlated bands to an automatic classifier to save computer time.

It is sometimes more advisable to use new and more suitable bands than the less correlated bands just mentioned. These are the *principal components* of the multi-band image which are certain linear combinations of the original bands, obtained by means of the Karhunen-Loève transformation [1], that are uncorrelated to each other. They correspond to the eigenvectors of the image covariance matrix C and are ordered by decreasing eigenvalue. The information content of the image is redistributed among these components in such a way that most of it is concentrated in the first few principal components, while the remaining components contain a minimum

amount of information (consisting mainly of noise) and can be disregarded without significant loss of information. In particular, in Landsat multispectral scanner (MSS) imagery the two most significant components usually contain more than 95 percent of the information in the original four bands. This is in agreement with the fact that the intrinsic dimensionality of those data is considered to be two [2].

The principal component technique is thus one of the best ways to select data to be assigned to the three fundamental colors to produce color displays of many-band images, or to be input to an automatic classifier. Nevertheless, it cannot be said *a priori* that the principal components of an image are exactly equivalent to the original bands for visual interpretation of all kinds of features. This is a consequence of the very subjective nature of human photointerpretation of images.

The principal components of an image must usually be contrast-enhanced in order to obtain displays suitable for photointerpretation, even though the variances of the most significant of them are greater than those of the original bands. However, when the variance of the displayed data is increased too much by means of a high contrast radiometric transformation, unwanted *artifacts* (i.e., apparent structures that do not really exist in the original data) can be introduced. These artifacts may be mistakenly considered to be real features by the photointerpreter.

Image principal components were first studied by Ready and Wintz [3] mainly for classification purposes. For some time little attention was paid to this and to the related technique of canonical analysis [4]. However,

Copyright 1978 by International Business Machines Corporation. Copying is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract may be used without further permission in computer-based and other information-service systems. Permission to *republish* other excerpts should be obtained from the Editor.

there has recently been a rekindling of interest [2, 5, 6] in these areas. This paper presents the first systematic study of the principal component technique for image display purposes. The method discussed constructs the principal components performing simultaneously an adequate linear radiometric enhancement without extra computational work. To achieve this, we do not simply use the coordinate transformation determined by the orthogonal matrix $\mathbf{G}$ that diagonalizes the image covariance matrix $\mathbf{C}$, but rather a more general coordinate transformation including both a dilatation and a change of coordinate origin. Furthermore, the same intensity level quantization interval is always used; thus, a finer level quantization is obtained when the contrast gain factor is higher. Consequently, displays with high data variances can be obtained completely free of artifacts. The price paid to achieve this goal is a slightly smoother appearance of the displayed image.

In the following sections the principal components of a general n -band digital image are defined, and a discussion is given of the most useful options for the simultaneous radiometric enhancement of these components. The signal-to-noise ratio (SNR) improvement is also considered, and the results are particularized to Landsat MSS images. The latter part of this paper is devoted to the geological study of Campo Arañuelo Basin, where we use Landsat MSS images and the technique described here. A summary is given of the most significant results of this work, with special emphasis on the usefulness of the principal components.

Determination of the principal components

Let us obtain the principal components of an n -band digital image whose intensity levels are m -bit integer numbers. We consider picture elements (pixels) as vectors $\mathbf{x}$ of the n -dimensional Euclidean space $\mathbf{R}^n$, whose coordinates in a given orthonormal base $\{\mathbf{e}_i, i = 1, \dots, n\}$ represent the pixel intensity levels in the n bands:

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i. \quad (1)$$

The digitization of intensities requires that pixels be particular vectors whose coordinates in the base $\{\mathbf{e}_i\}$ are restricted to be m -bit integers; i.e.,

$$x_i \in J_m \equiv \{\chi | \chi \in \mathbf{Z}, 0 \leq \chi \leq 2^m - 1\}, \quad i = 1, \dots, n. \quad (2)$$

In the following calculations, pixel intensities are allowed to take values on $\mathbf{R}$. Only in the final step will they be approximated by elements of J_m .

We also consider the n -band image as a vector random variable $\mathbf{X}$ on the space of events $\mathbf{R}^n$. Its coordinates in the base $\{\mathbf{e}_i\}$ are the random variables, $X_i, i = 1, \dots, n$, that correspond to the n original bands of the image:

$$\mathbf{X} = \sum_{i=1}^n X_i \mathbf{e}_i. \quad (3)$$

The covariance matrix $\mathbf{C}$ of the coordinates of $\mathbf{X}$ is defined as

$$\mathbf{C} = E\{[\mathbf{X} - E(\mathbf{X})] \otimes [\mathbf{X} - E(\mathbf{X})]\}, \quad (4)$$

where E stands for expectation and $\otimes$ denotes the tensor product. Since $\mathbf{C}$ is a symmetric matrix, it can be diagonalized. Let $\lambda_i, i = 1, \dots, n$, be the eigenvalues of $\mathbf{C}$ (i.e., the roots of the secular equation $|\mathbf{C} - \lambda \mathbf{I}| = 0$, in decreasing order: $\lambda_1 > \lambda_2 > \dots > \lambda_n$). We assume, for the sake of simplicity, that there are no degenerate eigenvalues, although this assumption is not essential. Let $\{\mathbf{g}_i, i = 1, \dots, n\}$ be the orthonormal base of the eigenvectors of $\mathbf{C}$ that satisfy

$$\mathbf{C} \mathbf{g}_i = \lambda_i \mathbf{g}_i, \quad (5)$$

and let $\mathbf{G}$ be the orthogonal matrix that diagonalizes $\mathbf{C}$:

$$\mathbf{G} \mathbf{C} \mathbf{G}^{-1} = \mathbf{\Lambda}, \quad (6)$$

where $\Lambda_{ij} = \lambda_i \delta_{ij}$ and $G_{ij} = (\mathbf{g}_i)_j, i, j = 1, \dots, n$.

The random variables

$$Y_i = \sum_{j=1}^n G_{ij} X_j, \quad i = 1, \dots, n, \quad (7)$$

are the coordinates of $\mathbf{X}$ in the base $\{\mathbf{g}_i\}$:

$$\mathbf{X} = \sum_{i=1}^n Y_i \mathbf{g}_i. \quad (8)$$

These variables are uncorrelated provided their covariance matrix is $\mathbf{\Lambda}$. Only if the probability distribution of $\mathbf{X}$ were normal would these new random variables be independent instead of being only uncorrelated. Consequently, the coordinates of a vector $\mathbf{x}$ in the base $\{\mathbf{g}_i\}$, i.e., the intensity levels of the corresponding pixel in the uncorrelated bands, are

$$y_i = \sum_{j=1}^n G_{ij} x_j, \quad i = 1, \dots, n. \quad (9)$$

These numbers do not belong to J_m . To approximate them by elements of J_m we introduce the function

$$F_{0,m}(\chi) = \begin{cases} 0 & \text{if } \chi \leq 0; \\ \text{Int}(\chi) & \text{if } 0 \leq \chi \leq 2^m - 1; \\ 2^m - 1 & \text{if } 2^m - 1 \leq \chi, \end{cases} \quad (10)$$

where Int stands for integer value. Then, the pixel intensities in the principal components of the image are finally defined as

$$z_i = F_{0,m} \left(a_i \sum_{j=1}^n G_{ij} x_j + b_i \right), \quad i = 1, \dots, n. \quad (11)$$

The parameters a_i and b_i are to be determined according to the statistical properties required for the principal component data. The elements of the covariance matrix Λ' of the principal components obtained by means of Eq. (11) are approximately equal to $a_i a_j \Lambda_{ij} = a_i^2 \lambda_i \delta_{ij}$, $i, j = 1, \dots, n$, the small actual differences being due to roundoff error and tail cutoffs introduced by the function $F_{0,m}$.

Radiometric enhancement options

Equation (11) allows us to radiometrically enhance the principal components while obtaining them, with no additional computation required. We have only to assign suitable values to the parameters $a_i, b_i, i = 1, \dots, n$. To see this, it suffices to realize that we can take $G'_{ij} = a_i G_{ij}$; hence, we have not increased the number of multiplications needed. Furthermore, some additive constant is always necessary in order to avoid the cutoff of all negative values of the uncorrelated bands y_i .

When the principal components are to be displayed either on a screen or by any kind of hard-copy recorder, the particular characteristics of the device and display method used (e.g., development of photographic film) must be considered in order to determine suitable values for the parameters $a_i, b_i, i = 1, \dots, n$. In general, one requires that the means of the statistical distributions of the principal components be a certain intensity level μ and that the most significant part of these distributions be included on an interval of intensity levels of width $2d$, centered at μ . Obviously, $[\mu - d, \mu + d] \subset J_m$. Then

$$\begin{aligned} b_i &= \mu - a_i E(Y_i) \\ &= \mu - a_i \sum_{j=1}^n G_{ij} E(X_j), \quad i = 1, \dots, n. \end{aligned} \quad (12)$$

For the determination of a_i a number of possibilities can be considered, the most simple and useful of which are the following.

Option 1

Take just

$$a_i = 1, \quad i = 1, \dots, n, \quad (13)$$

making no contrast enhancement. Sometimes the intensities of the original bands are not true m -bit numbers; however, there exists a number $m' < m$ such that $x_i \in J_{m'}$, although for some reason one prefers to think of these numbers as m -bit numbers. This option may then be appropriate. Otherwise, it may not be advisable since the lengths of the intervals on which the coordinates $y_i, i = 1, \dots, n$, take values are larger than those on which the coordinates $x_j, j = 1, \dots, n$, take values.

Option 2

To overcome the problem mentioned in option 1, divide by the maximum length increase factor, i.e., the ratio of

the length of the n -dimensional hypercube diagonal to the length of its side, and take

$$a_i = \frac{1}{\sqrt{n}}, \quad i = 1, \dots, n. \quad (14)$$

Option 3

One can apply to each principal component a radiometric enhancement such that the central part of its statistical distribution that includes ν standard deviations to each side of the mean value is assigned to the interval $[\mu - d, \mu + d]$. Then

$$a_i = \frac{d}{\nu \sqrt{\lambda_i}}, \quad i = 1, \dots, n. \quad (15)$$

Obviously, the ratios among the variances of the principal components are no longer maintained.

Option 4

The ratios among the variances of the principal components can be preserved by taking the same gain for all principal components:

$$a_i = \frac{d}{\nu \sqrt{\lambda_1}}, \quad i = 1, \dots, n. \quad (16)$$

In this way, the first principal component is radiometrically enhanced as in option 3, while the other principal components are enhanced proportionally.

The parameter ν , which determines the actual number of standard deviations assigned to an interval of length d in options 3 and 4, allows one to control the amount of contrast of the displayed data. Nevertheless, ν cannot be very large because of the image degradation inherent in the accumulation of pixels in intensity levels outside the interval $[\mu - d, \mu + d]$.

In addition to these options one can also replace some principal components with their negative images. This is equivalent to replacing the corresponding eigenvectors (e.g., $\mathbf{g}_k$) with their opposite vectors ($-\mathbf{g}_k$). It makes no difference because the phases of the eigenvectors are arbitrary. The negative of the k th principal component has intensity levels given by

$$z'_k = 2^m - 1 - z_k. \quad (17)$$

Thus, one obtains this negative in options 1 and 2 merely by the replacement

$$a_k \rightarrow -a_k, \quad b_k \rightarrow 2^m - 1 - b_k. \quad (18a)$$

The required replacements in options 3 and 4 are

$$a_k \rightarrow -a_k, \quad b_k \rightarrow 2\mu - b_k. \quad (18b)$$

In general, options 3 and 4 provide finer level quantization than options 1 and 2, as well as an artificial in-

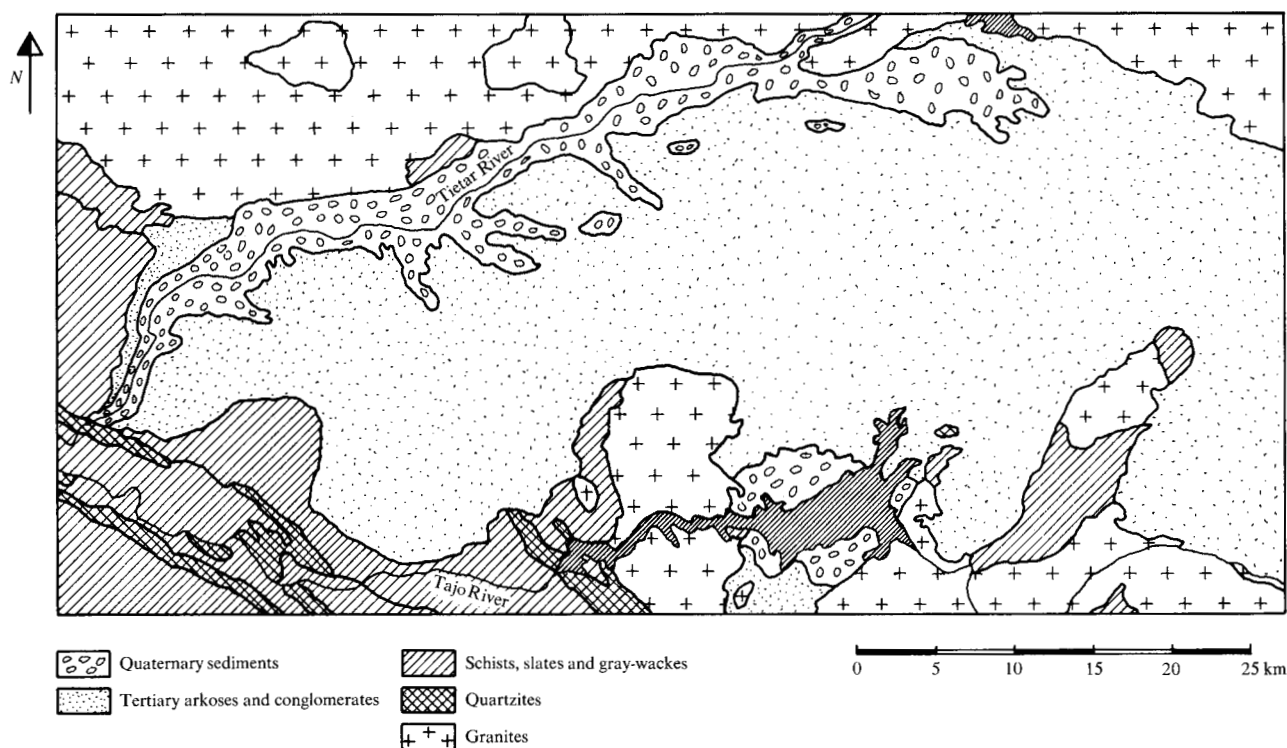


Figure 1 Lithologic scheme of Campo Arañuelo Basin as defined by the maps of the Instituto Geológico y Minero de España.

crease of variance. Displays produced by *a posteriori* radiometric enhancement of principal components obtained by means of options 1 and 2 usually appear rather harsh compared to those obtained by options 3 and 4. This is a consequence of the existence of empty intensity levels between consecutive occupied ones that can introduce unwanted artifacts when the difference between consecutive occupied levels is too large.

Signal-to-noise ratio improvement

Since the transformation matrix $\mathbf{G}$ is orthogonal, it is evident that $\text{tr } \mathbf{C} = \text{tr } \mathbf{A}$. That is,

$$\sum_{i=1}^n C_{ii} = \sum_{i=1}^n \lambda_i. \quad (19)$$

Thus, the reduction of a multispectral image to its principal components produces a redistribution of variance. If one defines the function

$$H(\mathbf{C}) = - \sum_{i=1}^n \frac{C_{ii}}{\text{tr } \mathbf{C}} \log \left(\frac{C_{ii}}{\text{tr } \mathbf{C}} \right), \quad (20)$$

it can be shown [7] that $H(\mathbf{C})$ is an entropy function whose minimum on the set of matrices connected with $\mathbf{C}$ by orthogonal transformations is $H(\mathbf{A})$, obtained when the principal components are used. The variances λ_i , $i = 1, \dots, n$, are thus a measure of the information contents of each component. By assuming that the distribu-

tion of the noise existing in the image data is symmetric in $\mathbf{R}^n$, it can be shown [3] that the improvement in the signal-to-noise ratio obtained by using the first principal component rather than the i th original band is

$$\Delta_i = \frac{\lambda_1}{C_{ii}}, \quad i = 1, \dots, n. \quad (21)$$

Application to Landsat images

We now particularize these results to the four-band images of the multispectral scanner (MSS) of Landsat satellites. We take $m = 8$, although Landsat computer-compatible tape data actually have seven bits in bands 4, 5 and 6, and six bits in band 7. This choice is dictated by the use of nine-track magnetic tapes for image storage and computers with eight-bit bytes for processing. Thus, option 1 can be used with no tail cutoff taking place.

The off-diagonal elements of the correlation matrix ρ , defined as

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, \quad (22)$$

usually have for these images values that range from 0.4 to 0.95 and higher. Actually, there are considerable variations in these values from image to image as well as in different subimages of the same image. In general, however, the most correlated pairs are the two visible (4 and

Table 1 Statistical parameters of the original bands.

Band	Mean	Standard deviation	Variance	I (bits/pixel)
4	15.13	3.81	14.52	3.80
5	22.05	5.78	33.41	4.40
6	36.53	9.08	82.45	5.05
7	20.41	5.66	32.04	4.45

5) and the two infrared bands (6 and 7), while bands 4 and 7 and 5 and 7 are usually the less correlated pairs.

The first two principal components (denoted as bands A and B) usually contain more than 95 percent of the whole image variance, while the other two (denoted as C and D) consist almost entirely of noise. In most cases bands A and B suffice for automatic classification (the intrinsic dimensionality of Landsat MSS data is considered to be two [2]) and for production of color prints for photo-interpretation.

The computer program used to obtain the principal components of a Landsat image, which is mostly written in IBM System/360 assembler language, uses a very fast algorithm to calculate the matrix *C*. The internal subroutine that accumulates the contribution of every pixel to the ten different elements of *C* uses only register-to-register instructions (apart from four "insert character" instructions) and needs only five multiply instructions to calculate the ten products. This algorithm is exact when one deals with Landsat images, but if the original data actually have eight significant bits there is a small error in five of the elements of *C* whose relative value is bounded to be less than 6×10^{-5} . For the construction of the principal components, this program uses a 16-Kbyte ($= 16 \times 1024$ bytes) look-up table that contains the values of $a_i G_{ij} x_j \times 16^4$ ($x_j = 0, \dots, 255$; $i, j = 1, 2, 3, 4$) as 32-bit integers.

When this program is run (either in batch mode or interactively by means of ER-MAN II [8]) on an IBM 360/65 computer it uses 58 Kbytes of main storage and takes 9 min 32 s of CPU time to compute *C* and *A* for a whole Landsat image (7.6×10^6 pixels $= 30.4 \times 10^6$ bytes of data), or 75.3 μ s/pixel. To construct the principal components using any of the discussed options, the program requires 92 Kbytes of main storage and an additional 11 min 36 s of CPU time, i.e., an additional 91.6 μ s/pixel.

Application to a geological problem

• Description of the area studied and expected goals

The Campo Arañuelo Tertiary basin, at the west end of the Tajo structural basin, includes part of the provinces of Toledo and Cáceres, Spain. Its center is Navalmoral de la Mata, about 160 km to the WSW of Madrid.

Two main lithologic groups (basement and sedimentary mantle) can be defined in this area with regard to the age of the rocks and their different dynamic behavior during the most recent movements [9]. The rocks forming the borders of the basin (see Fig. 1) are mainly granites and metamorphic rocks of Paleozoic age that have been fractured since Hercynian orogeny. In Alpine movements these rocks behaved like a rigid unit, and constituted the source areas that provided sediments for filling of the basin. These sediments are mainly arkoses and conglomerates of arkosic matrix, probably of Oligocene age. The activity of the numerous faults located in the basement-sedimentary mantle contact has enhanced the habitual topographic contrast between these materials. Among the fault and fracture trends existing in the basement rocks as a heritage of former orogenies, only those between the NNE-SSW and NE-SW trends clearly show the effects of a recent reactivation.

The expected goal of a global study of this area is to draw some conclusions about the geological history of the entire Tajo Basin in order to confirm the geological individuality of Campo Arañuelo Basin with respect to the rest of Tajo Basin in post-Oligocene times [9]. The existence of a crustal megastructure in central Spain constituted by an oval-shaped uplift with the major axis in the NE-SW direction, divided in two halves by an E-W fault zone, has been conjectured [10]. Accordingly, one would expect less activity during the Middle and Young Tertiary era in the southern half of this megastructure, in which Campo Arañuelo Basin is located, than in the northern half.

• Landsat imagery used

For the study of Campo Arañuelo Basin, four images of the multispectral scanners of Landsat 1 (formerly named ERTS-1) and Landsat 2 were initially considered. Two of them (scenes 2333-15404 of 21st December 1975 and 2350-10175 of 7th January 1976) were obtained in winter with low sun elevation angle (20° in both images). A preliminary analysis of the photographic data supplied by NASA indicated that snow covering the mountains of the northern boundary of the basin appeared saturated, rendering any interpretation of that area impossible. The topographic forms not covered by snow were enhanced by the shadows cast at that low sun elevation. These images provided useful geological information, given the correlation existing between geologic structures and landforms in this area, where dynamic processes have been rather recent. However, the poor illumination conditions made any discrimination among rock types by color differences very difficult. In a summer image (scene 2170-10201 of 11th July 1975) with high sun elevation angle (57°), rock type discrimination by color was easier, but the topographic forms were more difficult to see.

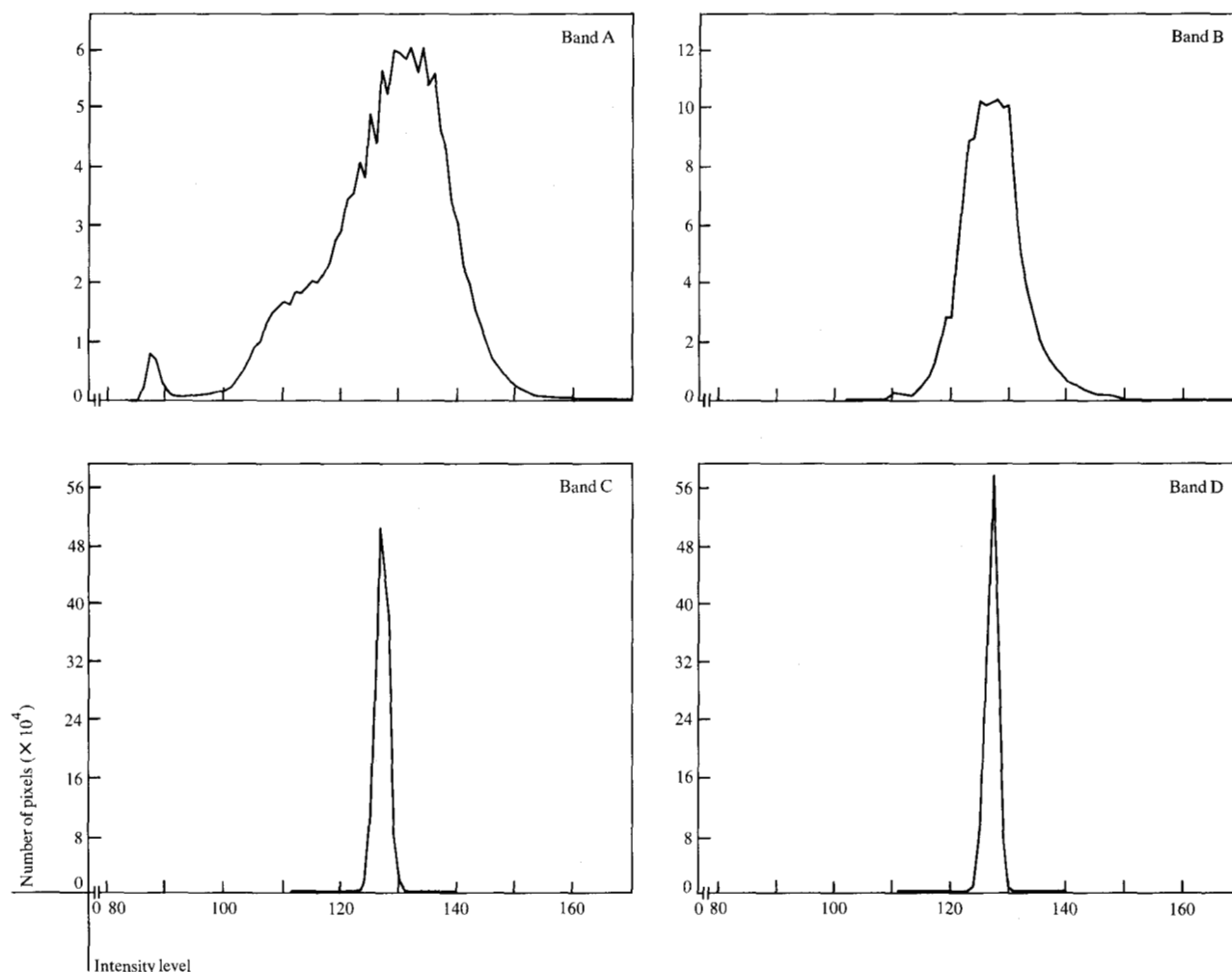


Figure 2 Histograms of the four principal components: abscissae represent intensity levels and ordinates are the number of pixels with each intensity. Notice the different ordinate scales.

It was decided that scene 1228-10325 of 8th March 1973 would be used because the absence of snow on this area and the intermediate sun elevation angle (37°) allowed good color discrimination of rock types, as well as a clear determination of topographic forms.

• Digital processing of the image

The method applied to geometrically correct the Landsat image [11] accounts for the systematic errors due to sensor characteristics and Earth rotation, as well as the effects of variations in satellite altitude and attitude. These variations are estimated by means of ground control points for which geographic and image coordinates are well established. Furthermore, the two functions that determine the mapping between input-image and output-image space account for the transformation of the image to Universal Transverse Mercator (UTM) projection. These functions are eventually approximated by two complete

polynomials of fifth degree in two variables which are actually used to make the image transformation. Data are interpolated by means of the nearest-neighbor algorithm [11] in such a way that output image pixels represent squares 50.8 m on a side.

Fourteen ground control points reasonably well distributed over the image were used. The accuracy of the geometric correction achieved by the method described can be estimated by the root mean square error in the location of the ground control points after the transformation. In our case this error was 131.5 m, satisfactory for our purposes.

From this corrected image a subimage (1560 pixels wide and 917 lines long) showing an area of $79.2 \times 46.6 \text{ km}^2$ in the provinces of Cáceres and Toledo (central Spain) was selected. It includes all Campo Arañuelo Basin. This is the image considered from now on. Table 1 shows the most significant parameters characterizing the

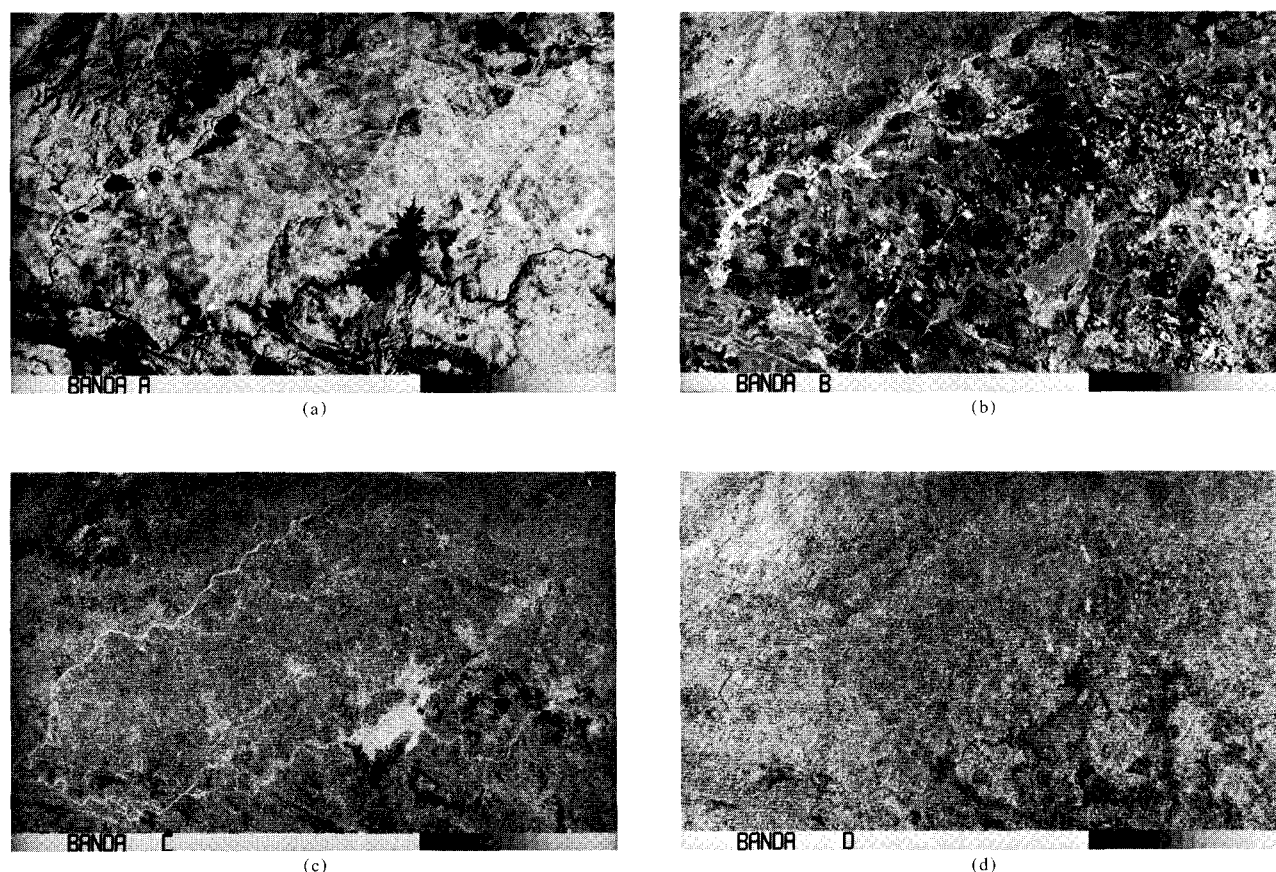


Figure 3 Principal components obtained by using option 3: (a) band A; (b) band B; (c) band C; and (d) band D.

statistical distribution of the pixel intensity levels of the four original bands. The information (listed in the last column in bits per pixel) is defined [12] by

$$I = - \sum_{k=0}^{255} p_k \log_2 p_k, \quad (23)$$

where p_k is the fraction of pixels with intensity level k .

The covariance matrix of this image, given by (4), is then

$$C = \begin{pmatrix} 14.51 & 20.54 & 22.19 & 10.87 \\ 20.54 & 33.43 & 30.89 & 14.40 \\ 22.19 & 30.89 & 82.38 & 49.40 \\ 10.87 & 14.40 & 49.40 & 32.04 \end{pmatrix}, \quad (24)$$

and the correlation matrix [Eq. (22)] is

$$\rho = \begin{pmatrix} 1.000 & 0.933 & 0.642 & 0.504 \\ 0.933 & 1.000 & 0.589 & 0.440 \\ 0.642 & 0.589 & 1.000 & 0.962 \\ 0.504 & 0.440 & 0.962 & 1.000 \end{pmatrix}. \quad (25)$$

The higher correlations are those between the two infrared bands (0.962) and the two visible bands (0.933),

while the correlation between one visible band and one infrared band is much smaller. Nevertheless, even the smallest correlation (0.440, between bands 5 and 7) is still rather high.

The eigenvalues of C (the nominal values of the variances of the principal components) are $\lambda_1 = 132.95$, $\lambda_2 = 27.05$, $\lambda_3 = 1.27$, and $\lambda_4 = 1.09$. The normalized eigenvectors of C (which indicate the directions of the principal components in $\mathbf{R}^4$) are

$$\begin{aligned} \mathbf{g}_1 &= \begin{pmatrix} 0.249 \\ 0.358 \\ 0.775 \\ 0.457 \end{pmatrix}; & \mathbf{g}_2 &= \begin{pmatrix} 0.443 \\ 0.770 \\ -0.285 \\ -0.361 \end{pmatrix}; \\ \mathbf{g}_3 &= \begin{pmatrix} 0.851 \\ -0.521 \\ 0.011 \\ -0.075 \end{pmatrix}; & \mathbf{g}_4 &= \begin{pmatrix} 0.135 \\ 0.092 \\ -0.564 \\ 0.809 \end{pmatrix}. \end{aligned} \quad (26)$$

Table 2 gives statistical information about the principal components obtained according to option 1 (no radio-

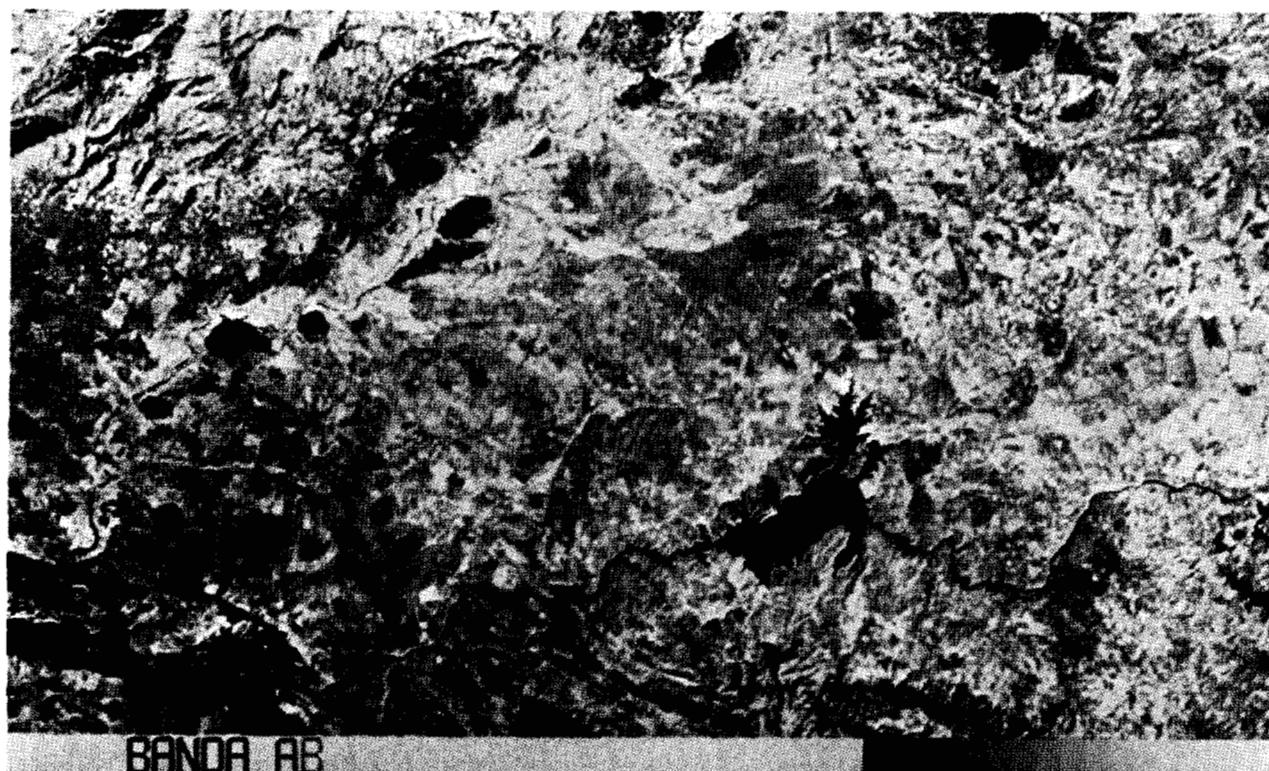


Figure 4 Two-color composite of band A (green) and band B (red) obtained by using option 3.

Table 2 Statistical parameters of principal components (option 1).

Band	Mean	Standard deviation	Variance	% of total variance	I (bits/pixel)	λ
A	127.00	11.53	133.01	81.9	5.45	132.95
B	127.00	5.21	27.13	16.6	4.36	27.05
C	127.00	1.17	1.36	0.8	2.23	1.27
D	127.00	1.08	1.17	0.7	2.06	1.09

metric enhancement). Notice that the differences between the actual and nominal variances (λ_i , $i = 1, 2, 3, 4$) due to roundoff errors are very small.

Bands A and B contain 98.5 percent of the total image variance, while the 1.5 percent contained in bands C and D is mainly due to noise. Thus, the first two principal components contain nearly all the information existing in the original four bands. These components constitute an excellent input to automatic classifiers, obviously better than the two less correlated bands, 5 and 7.

Histograms of the four principal components appear in Fig. 2 and illustrate the different information contents expected from bands A, B, C and D.

Equation (21) gives an estimate of the signal-to-noise ratio (SNR) improvement achieved by means of band A, although the most important noise component (the pecu-

liar striping produced by differences in calibration among the six sensors of each band that simultaneously provide the data of six consecutive image lines) is not spherically symmetric in R^4 . The SNR improvement with respect to band 6 (the one with highest variance) is 2.1 dB. This increases to 9.6, 6.0, and 6.2 dB with respect to the bands (4, 5, and 7, respectively) most often used to make false color displays.

• Photographic displays

It is obvious from Fig. 2 that even the first principal component obtained according to option 1 must be radiometrically enhanced *a posteriori* when it is to be displayed by means of a device with a dynamic range $2d$ much greater than $2\sqrt{\lambda_1}$. On the other hand, one can also use options 3 or 4 to construct the principal components

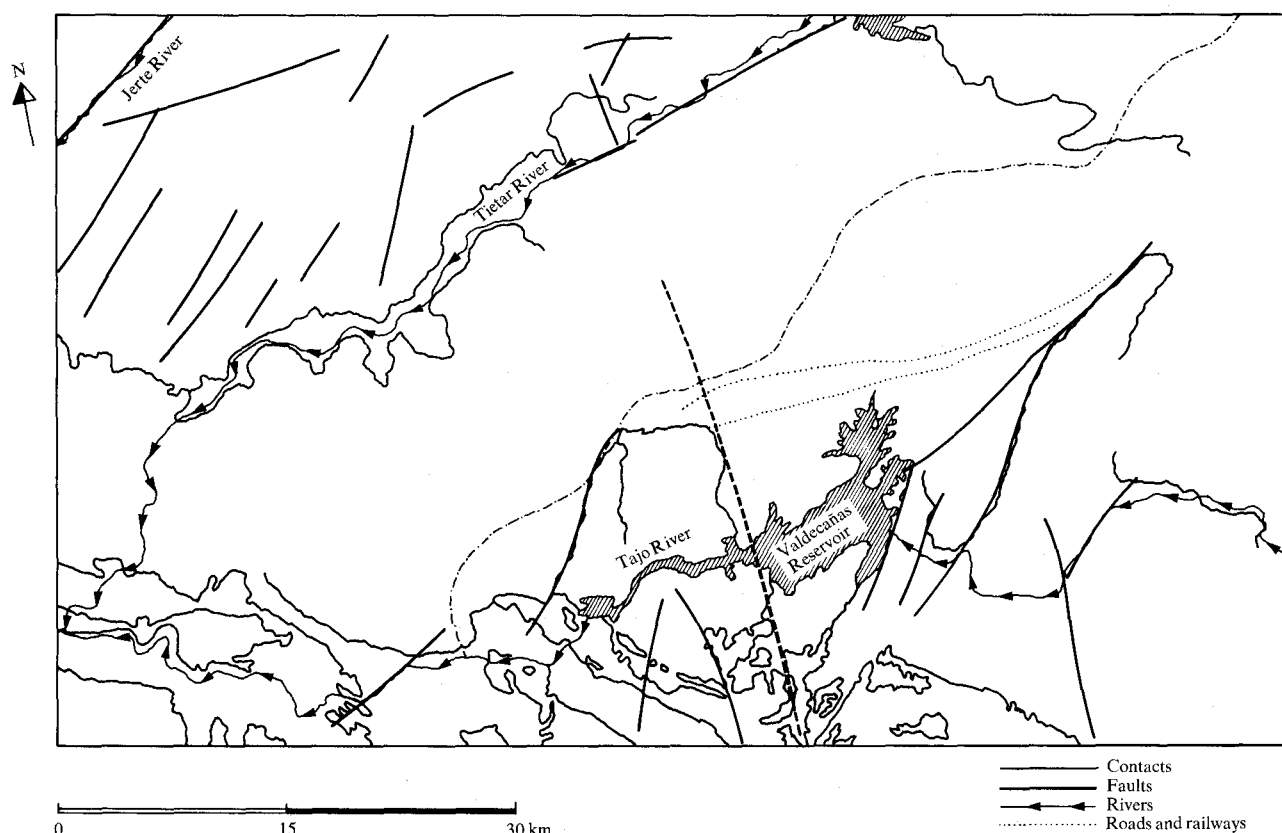


Figure 5 Geologic map of Campo Arañuelo Basin obtained by photointerpretation of band A.

already enhanced. We use option 3; hence the smaller the eigenvalue λ_i of a principal component, the greater the contrast gain a_i that we take for it.

A DICOMED D-47 film recorder (Dicomed Corp., Minneapolis, MN) was used to display the images. This recorder can distinguish 256 intensity levels (from 0 to 255) and d can be taken as 127.5. Good prints can be obtained with $\mu = 255/2 = 127.5$, and ν between 2 and 3. Conventionally, $\nu = 2.65$ is taken because the tail cutoffs produced in a Gaussian intensity level distribution accumulate a fraction $1/256$ of the pixels in each of the end levels (0 and 255). The gain factors used were $a_1 = 4.17$, $a_2 = 9.25$, $a_3 = 42.66$, and $a_4 = 46.13$. Figures 3(a-d) show the principal components obtained in this way. A gray wedge with 17 intensity levels (0, 16, 32, ..., 240, and 255) was added to the bottom of these images for use as a gray level reference. The wedge also provides a scale reference, its length corresponding to 25.9 km on the image.

When looking at Figs. 3(a-d), keep in mind the different contrast gain factors used. As pointed out previously, only bands A and B are useful for photointerpretation. Band A of our image was very suitable for visualizing structural features and, in general, for studying geology, while band B was more suitable for studying agriculture

and forestry. Note that the band B contrast has been increased 2.22 times over that of band A. The contrasts of bands C and D have been increased 10.23 and 11.06 times over that of band A to demonstrate that these bands contain almost solely noise, striping being the dominant component. When option 4 is used (the same contrast gain being used for all principal components), bands C and D appear uniformly gray and display no information at all.

Bands A and B can be used to produce two-color composite displays. In Fig. 4, band A is displayed in green, band B in red.

• Photointerpretation

Prior to any Landsat data analysis, a regional survey was made that included field rock identification, fault and fracture measurement, and mapping. Geological features attributable to older-than-Tertiary processes in the basement rocks were ignored since our interests were confined to the recent geological history of the area (as deduced from studies of the sediments, their facies and basement-cover spatial relationships [9]).

The dynamic conditions deduced corresponded to a pulsational rise in the old-faulted blocks in the source areas, this reactivation beginning in the older Oligocene

and lasting until the end of this period. These minor preliminary movements terminated with a major dynamic phase that ended the sedimentary process and moved the old-generated reverse faults located in the basement limit. Thus, the Paleozoic materials ended up on top of the Tertiary layers just deposited. It can be assumed that this marked the end of the Tertiary dynamic history of this area, while in the rest of the Tajo Basin the dynamic process extended further in time to produce a complete Miocene series. The eastern limit of Campo Arañuelo Basin is determined by a basement-faulted block in the area of Talavera de la Reina that gives rise to a step in the deep sedimentary basin floor. This step is detectable by means of stratigraphic and hydrogeological research and is also noticeable in Landsat images.

Landsat imagery was helpful in the global study of Campo Arañuelo Basin by facilitating rock-unit mapping, and also through its capacity for displaying major geological features such as regional scale faults, folds, and basin boundaries. The great ability of band A to differentiate materials, to visualize faulted structures, and to determine topographic landforms is demonstrated by its photointerpretation (Fig. 5). The basement-sedimentary mantle contact is easily followed in the northern boundary, due both to the thick Quaternary sediments giving rise to fertile irrigated lands that are well differentiated on Landsat imagery, and to the important basement topographic elevation in this border. These basement rocks are mainly granites, and slight color differences on the image involve petrological differences in the rocks. In the southern boundary the basement-mantle contact appears very clear when it is related to structural or topographic features. Among the rocks that constitute this border, the SE-NW folds in the southeastern corner of the image are clearly visualized not only in their general trend but also in smaller details because rock layer differentiation is possible. However, the contact delineation between granite and metamorphic rocks inside the basement becomes more difficult because these two kinds of rocks do not give rise to contrasted landforms; they show denudation uniformity. However, the quartzite-schists layers in the folds show a differential erosion that is very easy to detect on Landsat images. In the sedimentary cover, a Tertiary facies classification is not possible because of the qualities inherent in the sediments. This classification cannot be achieved through high-altitude aircraft photography and is difficult even in field work. Faults and fractures affecting granites in the northern border appear very clear, as expected for such an area, where features involving a youthful stage of geomorphological evolution are frequent. In the southern border the NE-SW reverse faults conforming the basement-mantle contact can be easily drawn all along their paths. Other NE-SW fractures and faults are seen in this area; most of these are

either conforming Paleozoic rock contacts or displacing folded layers. In some cases, it is possible to observe slight lineations in the Tertiary cover that reflect underlying basement fractures.

Summary

This article reports the first systematic study of the principal components of an n -band image specifically intended for photointerpretation purposes. The method for obtaining the principal components allows one to perform different linear radiometric enhancements that can avoid artifact creation by means of a finer intensity level quantization. Furthermore, these radiometric enhancements do not require any additional computational work. Fast algorithms are used to obtain the principal components; the one used to calculate the covariance matrix of a Landsat image is especially fast.

In the Landsat image studied the first two principal components contain nearly all the information existing in the four original bands; they contain 98.5 percent of the total image variance. Therefore, two-color composites of these two components are approximately equivalent to the four original bands; as such, they are slightly better, from the point of view of information content, than any three-color composite of the four original bands. Obviously, the improvement achieved by means of three-color composites of the first principal components, with respect to any other color display of the original bands, increases dramatically as the number of bands is increased.

The global insight into Campo Arañuelo Basin that has been achieved by means of Landsat imagery could have been obtained neither from field work nor from interpretation of the hundreds of aircraft photographs needed to cover the area. Furthermore, the use of the first principal components (which proved to be appropriate for geological studies) permitted simultaneous visualization of information that was scattered in the original bands. In fact, band A is mainly constructed from bands 6 and 7. Individually, band 6 emphasizes geomorphologic features, while band 7 emphasizes geostructural features [13], but band A shows both features equally well.

As a consequence of this work we have confirmed the previous ideas about the geological individuality of Campo Arañuelo Basin with respect to Tajo Basin in later Tertiary times. This is in agreement with the conjecture of a crustal megastructure in central Spain [10], in which different dynamic behavior in the northern and southern halves would imply a shorter geological history in Campo Arañuelo Basin.

Acknowledgments

This work is the result of a collaboration between the IBM Madrid Scientific Center and the Department of In-

ternal Geodynamics of the Universidad Complutense de Madrid, and is part of a larger geological study of Tajo Basin that is being conducted by this University. The authors thank the Instituto Geográfico Nacional for photographic recording of image data, and the members of the Applications Department of the IBM Madrid Scientific Center for discussions and comments; in particular, J. González and F. P. Palóu for programming assistance.

References

1. M. Loève, *Probability Theory*, D. Van Nostrand Co., Inc., Princeton, NJ, 1955.
2. S. G. Wheeler, P. N. Misra, and Q. A. Holmes, "Linear Dimensionality of LANDSAT Agricultural Data with Implication for Classification," *Proceedings of the 3rd Machine Processing of Remotely Sensed Data Symposium*, Purdue University, W. Lafayette, IN, 1, 2A-1 (1976).
3. P. J. Ready and P. A. Wintz, "Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery," *IEEE Trans. Commun.* COM-21, 1123 (1973).
4. (a) M. M. Taylor, "Principal Components Colour Display of ERTS Images," *Proceedings of the 3rd Earth Resources Technology Satellite-1 Symposium*, NASA, Goddard Space Flight Center, Washington, DC, 1B, 1877 (1973).
(b) A. Fontanel, C. Blanchet, and C. Lallemand, "Enhancement of Landsat Imagery by Combination of Multispectral Classification and Principal Component Analysis," *Proceedings of the NASA Earth Resources Survey Symposium*, NASA, Johnson Space Center, Houston, TX, 1B, 991 (1975).
5. (a) G. J. McMurtry and F. Y. Borden, "Principal Components and Canonical Analysis for Skylab Channel Evaluation," *Proceedings of the 4th Machine Processing of Remotely Sensed Data Symposium*, Purdue University, W. Lafayette, IN, 1, 77 (1977).
(b) G. D. Lodwick, "An Analysis of ERTS Data in its Role in Terrain Classification," *Proceedings of the International Conference on Computer Application in Developing Countries*, Asian Inst. of Technology, Bangkok, Thailand, 1, 653 (1977).
6. A. Santisteban and L. Muñoz, "Application of Image Principal Component Technique to the Geological Study of a Structural Basin in Central Spain," *Proceedings of the 4th Machine Processing of Remotely Sensed Data Symposium*, Purdue University, W. Lafayette, IN, 1, 228 (1977).
7. S. Wantanabe, "The Loève-Karhunen Expansion as a Means of Information Compression for Classification of Continuous Signals," *Technical Report AMRL-TR-65-114*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1965.
8. "Earth Resources Management (ER-MAN II). User's Guide," IBM SB11-5008-0, IBM European Publication Center, Copenhagen, Denmark, 1976.
9. L. Muñoz, "Relación Basamento-Cobertera en un Segmento del Límite Meridional de la Depresión de Campo Arañuelo," Universidad Complutense de Madrid, Spain, 1976 (unpublished report).
10. M. Alía, "Una Megaestructura de la Meseta Ibérica: La Bóveda Castellano-Extremeña," *Estudios Geológicos* 32, 229 (1976).
11. R. Bernstein, "Digital Image Processing of Earth Observation Sensor Data," *IBM J. Res. Develop.* 20, 40 (1976).
12. L. Brillouin, *Science and Information Theory*, Academic Press, Inc., New York, 1971.
13. L. Muñoz and A. Santisteban, "Estudio Geológico de la Depresión de Campo Arañuelo por medio de Imágenes Landsat," *Memorias del Seminario Internacional sobre el Uso de los Sensores Remotos en el Desarrollo de los Países*, IBM Report CCAL 77-06, Centro Científico de América Latina, México D.F., México, 1977, p. 135.

Received September 15, 1977; revised December 19, 1977

A. Santisteban is with the IBM Madrid Scientific Center, Paseo de la Castellana 4, Madrid 1, Spain; L. Muñoz is with the Geodynamics Department of Universidad Complutense, Madrid 3, Spain.